NAG

momentum

$$w_{t+1} = w_t - v_t$$

where (past velocity)

$$v_t = \beta v_{t-1} + \eta \nabla w_t$$

(∇w_t → gradient at that point.)

$\beta \rightarrow$ decay factor.

$$w_{t+1} = w_t - \eta \nabla w_t$$ ← old gradient descent

NAG

∴ The major difference between Nesterov Accelerated gradient & mom.

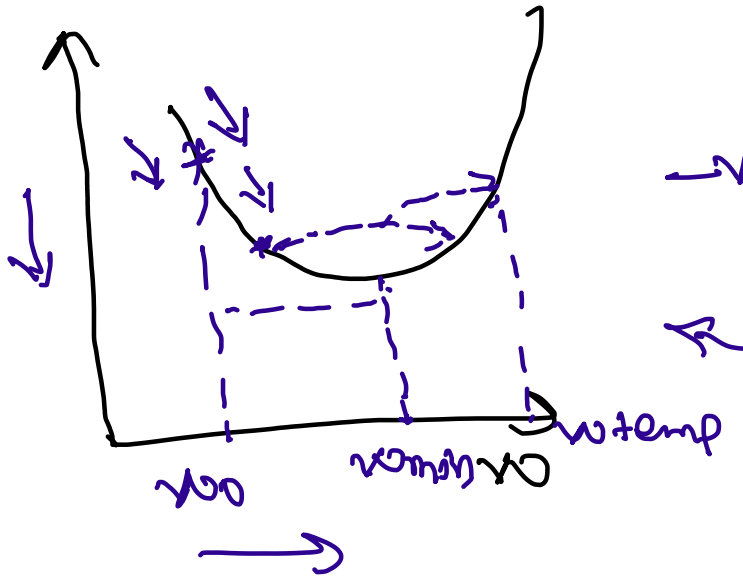
History of velocity + gradient at that point.

↓

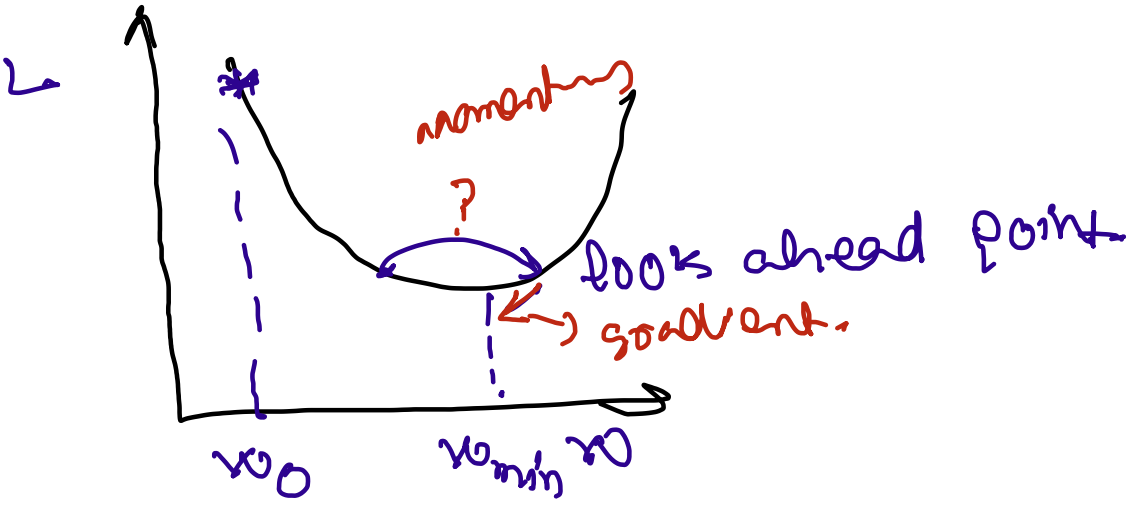
In momentum we take a jump together.

∴ The only different with what is after taking jump on velocity - the point after landing is considered and again as jump

Geometric Intuition



NAG



Major Disadvantage

NAG → oscillation → Dampen.

Keras Code

24 July 2022 13:18

```
tf.keras.optimizers.SGD(
    learning_rate=0.01, momentum=0.0, nesterov=False, name="SGD", **kwargs
)
```

To implement SGD

momentum
- momentum = 0-9
- nesterov = false

NAG
- momentum = 0-9
- nesterov = True